特集記事

AI 深層学習を用いた作業安全監視技術

東京大学大学院工学系研究科原子力専攻
出町　和之　Kazuyuki DEMACHI

1．はじめに

我が国の平成 26 年～ 30 年の労働災害死傷者数は年平均で約 12 万人にも達し、そのうち年平均死亡者数は 900 人を超える。業種別にみると、平成 30 年で死亡者数が最も多い 3 業種は順に、建設業、製造業、第三次産業、死傷者数が最も多い 3 業種は順に、第三次産業、製造業、陸上貨物運送事業であった。事故の型別発生状況では、墜落・転落が死亡事故者数の約 3 割を占め、次いで交通事故が 2 割、はさまれ・巻き込まれが約 1 割と続く。厚生労働省は「平成 30 年度～令和 4 年度の第 13 次労働災害防止計画にて死亡災害の 15% 以上減少、死傷災害の 5% 以上減少」という数値目標を設定している。

事業者は労働安全のために自主的な安全衛生活動やリスクアセスメント活動など多大な取り組みをしている筈であるが、現状では労働災害発生ゼロは実現できていない。その理由のひとつが、労働者や監督者のいわゆるヒューマンエラーと呼ばれるうっかりミスや失敗などである。ヒューマンエラーはどんなに意識していても完全に防ぐことは不可能であり、一定の確率で発生してしまう。

一方で、近年になって深層学習はさまざまな分野で大成功を収めている。自然言語処理（自動会話や文書校正）、機械との融合（二足歩行ロボット、自動倉庫）、医療・薬学応用（病理診断、医師国家試験合格）、物体・画像認識（無人コンビニ、顔認証、人物認証）、自動運転（運転距離 10,000,000 マイル到達）、音声解析（議事録の自動テキスト化）、脳波利用（脳波でゲーム操作）、ゲーム攻略（人間レベルを超えるゲーマー）などなど、その応用範囲の広さはまさに驚異的である。

中でも特に画像データは多くの情報を含むため、建設現場や工場などにおける作業者の危険状況を検知する技術への、深層学習（AI）の応用が期待されている。しかし現存する画像 AI を用いた危険検知は、「通常 / 逸脱」の二値判定もしくは画像識別の組み合わせに対し予め○×を設定しているに過ぎず、状況の組み合わせが無数にあり得る作業安全監視への適用は困難である。一方、原子力事業所などでは労働安全に限らず原子炉安全、核セキュリティなどのための行動ルールが精緻に文書化されており、これらルール文書に則って撮影された画像情報を判定することが望ましい。すなわち、撮影された画像情報がルールに違反しているかもしくはルールに合致しているかを柔軟に判定できれば実用性は格段に高くなるはずである。しかしながらそのためには、画像 AI と言語 AI という異なる AI 同士のインターフェイスが必要となる。

前述のように現在、多くの AI 研究が行われている。しかしその殆どは各々の分野の中だけで閉じており、相互乗り入れ、すなわちインターフェイスは皆無である。一方、人間の脳は言語、視覚、聴覚、触覚などあらゆる認識情報を連携させた処理を普通に行っている。もし異種 AI 同士が脳のような相関が可能となれば、AI の能力は新展開を迎えるはずであるが、これを困難にしている理由の一つが異種 AI 間に共通するデータ形態が不在なことであろう。

では共通データ形態として何があり得るのか？東京大学・出町研究室では、比較台帳、決定木、文ベクトル分散表現、グラフ構造化の 4 候補の検証を経たのち、グラフ構造が最適であるとの結論に至った。グラフ構造とはグラフ理論とも呼ばれ、モノやヒト同士の関係性を点（ノード）と矢印（エッジ）の連結で表現する手法である。本研究では、グラフ構造を共通データ形態とする画像 AI と自然言語処理 AI のインターフェイス開拓を目的に、3 つのアルゴリズムを開発した。これにより、例えば画像に映る状況の危険・安全を、判定基準となるルール文を入力するだけで自動で判定できるなど、非常に利便性と汎用性の高い技術の実装が可能となる。

アルゴリズム #A：画像情報のグラフ構造化
アルゴリズム #B：文情報の階層型グラフ構造化
アルゴリズム #C：グラフ構造の比較による判定

これらのアルゴリズムを実装したプログラムによるデモ動画の検証では 90%以上の危険判定精度が得られた。本研究により、画像と文とのインターフェイスを実現するための基盤技術は確立されたと言える。

2．画像 AI と言語 AI のインターフェイス

2.1 アルゴリズム #A ：画像情報のグラフ構造化

(1) 物体認識

画像情報をグラフ構造化するためには、まず深層学習による物体認識が必要である。物体認識には YOLOv3[1] を採用した（図 1）。YOLOv3 とは、画像に写る物体の座標を推測しかつ物体が何であるかを識別する深層学習モデルである。さらに YOLOv3 の特徴量抽出部（Darknet-53）を MobileNet[2] に置換することでリアルタイムの物体検出モデルに改良した（FPS=11, 1 秒間に 11 ～ 12 回）。図 2 に、YOLOv3 を用いて人物（Person）、椅子（Chair）、モニタ（tvmonitor）、ノート PC（Laptop）、キーボード（keyboard）、マウス（mouse）、本（book）を認識した結果の例を示す。枠内の数字は識別の確度を表す。

(2) 動作認識

姿勢認識には 2 次元の姿勢認識には XNect[3] を使用した。XNect とは画像に映る人物の 3 次元モーションを識別する深層学習モデルである。図 3 に XNect のアーキテクチャを、図 4 に XNect による多人数 3D 姿勢認識の結果の例を示す [3]。同様の技術に Microsoft 社 Kinect によるモーションキャプチャがあるが、Kinect は赤外線レーザーによる深度センサを使用するのに対し XNect は普通の RGB カメラ画像からの認識が可能である上、さらにステレオカメラのような複数画像も必要とせず、高い汎用性を有する。

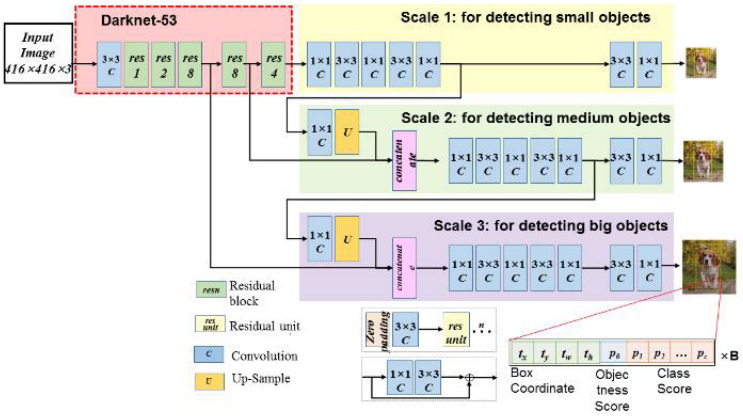

図 1 YOLOv3 のアーキテクチャ [1]

図 2 改良後の YOLOv3 による物体認識結果

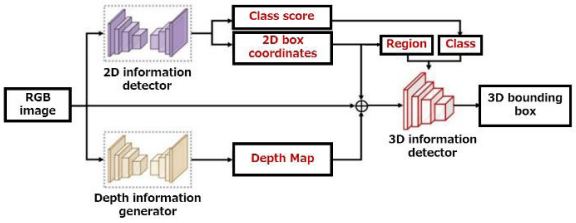

図 3 XNect のアーキテクチャ [3]

図 4 Xnect によるモーションキャプチャ [3]

画像に映る状況（Scene）を識別する手法に Scene Graph[4] がある。これは、画像と物体ー人物間の相関関係で 1 つのセットとなる教師データを大量に深層学習モデルに学習させ、画像を入力として物体ー人物相関を出力する手法である。しかし、Scene Graph は学習に長時間を要しかつ学習済みの状況のみ推定することから本研究ではこれを採用せず、YOLOv3 と XNect が検出した物体 S_i と人体部位 T_j の座標から求まる相互ユークリッド距離に基づき関係性を簡易的に推定する手法を提案した（図 5）。

すなわち、図 5 の赤数字のようにユーク

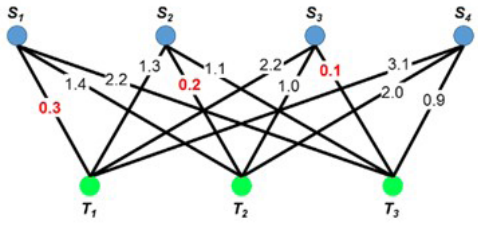

図 5 ユークリッド距離に基づく関係性抽出

リッド距離が最小かつ閾値以下となる物体 S_i と人体部位 T_j とが「相関あり」と推定した。この相関を点（ノード）と矢印（エッジ）の連結で表現することで、画像グラフ構造（図 6）が得られる。すなわち、画像グラフ構造は画像に映る状況の「要約」であるとも言える。

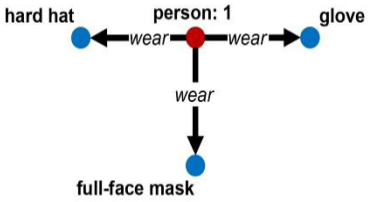

図 6 画像グラフ構造の出力結果の例

2.2 アルゴリズム #B ：文情報のグラフ構造化

このアルゴリズムでは、建設現場などにおける作業安全上のルール文書を対象にしてグラフ構造化モデルを構築した。ルール文書は、禁止ルール文、遵守ルール文、その他の文と、おおまかに 3 種類の文で構成される。本研究では、作業安全判定に必要な禁止・遵守ルール文の自動抽出を、自然言語処理 AI 手法である BERT（Bidirectional Encoder Representations from Transformers）[5] により行った。BERT は二値分類に特に高い性能を発するため、1 段階目：禁止・遵守ルール文とその他の文、2 段階目：禁止ルール文と遵守ルール文の 2 段階分類を用いた。

抽出された禁止ルール文、遵守ルール文に係り受け解析 [6] による形態素分析、記号化、品詞タグ付けを行うことで、文を構成する形態素間の依存関係が得られる。これにオントロジー解析 [7] を適用することにより依存関係を論理表現に変換した。

さらにこれを可視化してグラフ構造化したものが、出町研で提案する文グラフ構造である（図 7）。この図はある人物（person A）に対する 5 つの危険判定文をグラフ構造とした場合を模擬したものである。ここで ITMN のノード（点）を持つグラフ構造は”If ○○ then must not ××”、すなわち「○○の場合は××であってはならない」という禁止ルール文に相当し、ITM のノードを持つグラフ構造は”If ○○ then must ××”、すなわち「○○の場合は××でなければならない」という遵守ルールに相当する。ITMN, ITM のノードにはそれぞれ○○や××に相当する特徴ノードがエッジにより接続されており、二股のエッジは OR（または）を意味する。

ITMN ノードは禁止ルールを表すため、これに接続されている特徴ノードのすべてが判定対象の画像グラフ構造に含まれる場合、この禁止ルールに違反した状況（=Bad）を画像から識別できたと判定する。一方で、判定対象の画像グラフ構造の特徴ノードのすべてが遵守ルールを表す ITM ノードに接続される特徴ノードに含まれる場合、この遵守ルールを履行している状況（=Good）を画像から識別できたと判定する。

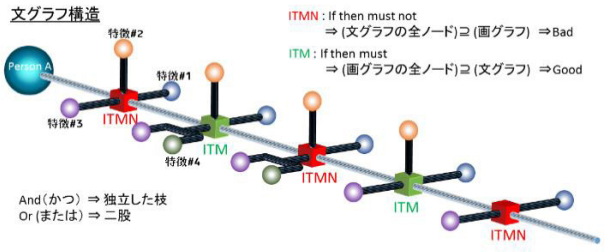

図 7 提案した文グラフ構造の例

2.3 アルゴリズム #C ：グラフ構造の比較による判定

このアルゴリズムでは、前述のように画像グラフ構造と文グラフ構造との比較により禁止ルールの違反、遵守ルールの履行を自動判定する。図 8 のように、自然言語処理 AI にてあらかじめ ITMN と ITM とに分けて展開した文グラフ構造を準備しておく。一方、画像 AI 側ではリアルタイムで画像グラフ構造を作成し、文グラフ構造の各 ITMN ルールおよび ITM ルールとの比較で Good, Bad を判定する。

例えば上側の画像グラフ構造の 3 つの特徴ノードは、1 つめの ITMN ノードの全ての特徴ノードを含むために

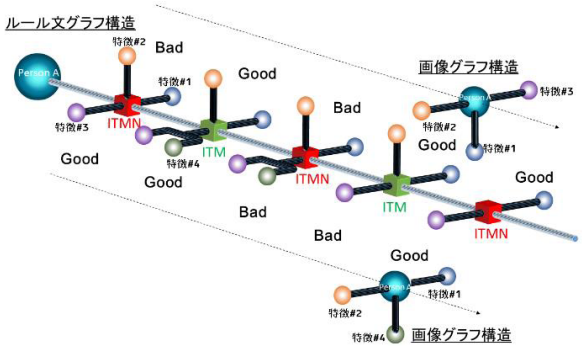

図 8 文グラフ構造と画像グラフ構造の比較による判定

Bad と判定され、２つめの ITM ノードの持つ特徴ノードの含まれるため Good と判定される。

３．結果

以上の３つのアルゴリズムを用い、表 1 に示す作業安全上の 7 種類のルールを例として、開発したアルゴリズムによる危険検知・判定性能の評価を行った。判定精度は、Precision（精度）および Recall（再現率）により評価した。

Precision（適合率）= TP/（TP+FP）　　(1)

Recall（再現率）= TP/（TP+FN）　　(2)

ここで TP は正解が真（True）であるときに推定結果も真（Positive）である割合、FP は正解が偽（False）であるときに推定結果が真（Positive）である割合、FN は正解が偽（False）であるときに推定結果も偽（Negative）である割合である。

表 1 検知判定対象とした作業安全上の 7 種のルール

1	建設現場ではヘルメットを着用すること
2	廃炉現場では全面マスクを着用すること
3	建設現場では防塵マスクを着用すること
4	グラインダーを操作するときは 保護メガネを装着すること
5	高所で作業するときは フルハーネスを着用すること
6	建設現場では手袋を着用すること
7	グラインダーを操作するときは 両手を使うこと

図 9 に、人物（person）がヘルメット（hart hat）を装着（wear）している画像からの、アルゴリズム #A による物体および姿勢認識結果の例を、図 10 に図 9 の結果からの関係性抽出結果の例を示す。このように複数の人物がいる場合でも個別に関係性を正しく抽出できた。

図 11 に”Wear a hard hat on a construction site.（建設現場ではヘルメットを着用すること）“という表 1 の 1 番目の条件付き禁止ルール文の、アルゴリズム #B による依存関係抽出結果を示す。

また図 12 に、表 1 の 7 種のルールをアルゴリズム #B により論理表現化した結果の例を示す。

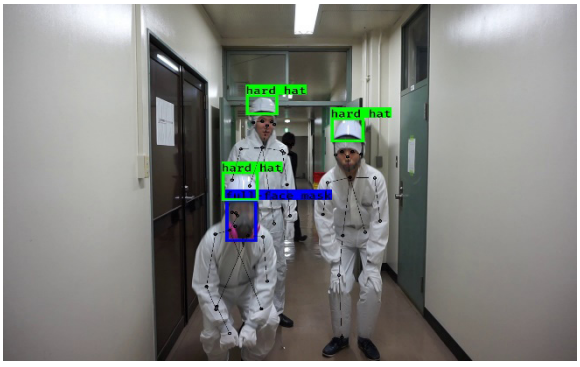

図 9 アルゴリズム #A による物体・姿勢認識

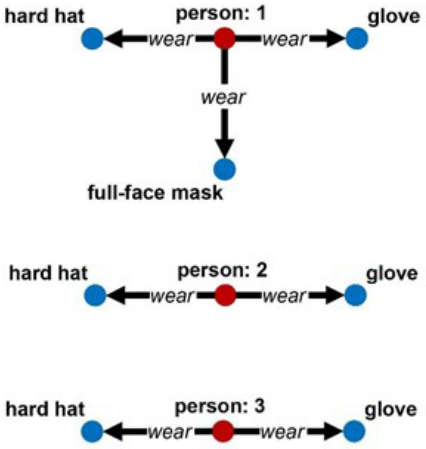

図 10 画像からの関係性抽出の例

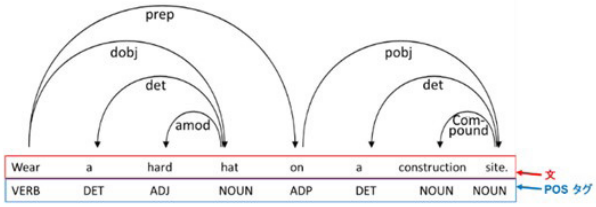

図 11 依存関係ツリー抽出結果の例

1)

$((person, on, construction\ site)) \rightarrow$
$(\square) \vee ((person, wear, hard\ hat))$

2)

$((person, on, decommisioning\ site)) \rightarrow$
$(\square) \vee ((person, wear, hard\ hat))$

3)

$((person, on, construction\ site)) \rightarrow$
$(\square) \vee ((person, wear, dust\ mask))$

4)

$((person, operate\ grinder)) \rightarrow$
$(\square) \vee ((person, wear, safety\ glasses))$

5)

$((person, on, height)) \rightarrow$
$(\square) \vee ((person, use, body\ harness))$

6)

$((person, on, construction\ site)) \rightarrow$
$(\square) \vee ((person, wear, glove))$

7)

$((person, operate, grinder)) \rightarrow$
$(\square) \vee ((person, use, two\ hands))$

図 12 条件付き禁止・遵守ルールの論理表現の例

図 13 に表 1 の 7 種の危険状況の、アルゴリズム #C による検知・判定精度の結果を示す。横軸はカメラから被写体までの距離（m）、縦軸は Precision（適合率、%）、Recall（再現率、%）を示す。7 種のルールに対する検知・判定精度の平均は、適合率 =94.2%、再現率 =84.4% と高精度判定の結果が得られた。なおこのとき計算速度は CPU: Intel Core i7-7820X(8, 3.6GHz), GPU: NVIDIA GeForce GTX 1080Ti/11GB で 7.95FPS であり、リアルタイム判定のためには十分に高速である。

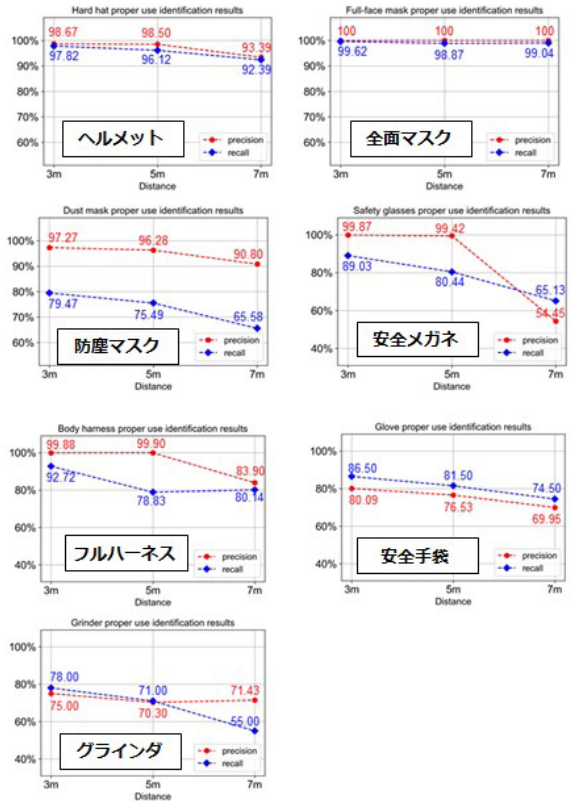

図 13　7 種の作業安全上の危険状況の判定精度

4．まとめ

画像 AI と自然言語処理 AI とを、グラフ構造という共通のデータ形態を介して相関させる異種 AI インターフェイスを提案し、これを実現するための 3 種のアルゴリズムを開発するとともに、それぞれ自動計算用に実装した。

作業安全上の 7 種の危険状況検知による検証では、適合率・再現率とも高精度判定の結果が得られた。これにより異種 AI インターフェイス手法の基盤は確立できたと考える。

本アルゴリズムの最大の利点は、利用者は検知したい禁止・遵守状況を「ルール文」として直接入力するだけでよく、利便性が各段に高いことである。

今後は、検知対象となる危険・遵守状況を広範囲に拡張するとともに、さらに画像と自然言語処理以外の異種 AI 同士のインターフェイスへの展開を図る。

参考文献

[1]Redmon, J., and Farhadi A., "Yolov3: An incremental improvement." arXiv preprint arXiv:1804.02767 (2018).

[2]Sanjay Kumar K.K.R., Subramani G., Thangavel S., Parameswaran L. (2021) A Mobile-Based Framework for Detecting Objects Using SSD-MobileNet in Indoor Environment. In: Peter J., Fernandes S., Alavi A. (eds) Intelligence in Big Data Technologies−Beyond the Hype. Advances in Intelligent Systems and Computing, vol 1167. Springer, Singapore.
Fhttp://doi-org-443.webvpn.fjmu.edu.cn/10.1007/978-981-15-5285-4_6

[3]Mehta,.D. et al., “Xnect: Real-time multi-person 3d human pose estimation with a single camera.”, arXiv preprint arXiv: 1907.00837 (2019)

[4]Danfei Xu, Yuke Zhu, Christopher B. Choy and Li Fei, “Scene graph generation by iterative message passing”, Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2017)

[5]Jacob Devlin, Ming-Wei Chang, Kenton Lee and Kristina Toutanova, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding”, ERT(Bidirectional Encoder Representations from Transformers), arXiv preprint arXive: 1810.04805 (2018)

[6] 松田寛 , 大村舞 , 浅原正幸 . 短単位品詞の用法曖昧性解決と依存関係ラベリングの同時学習 , 言語処理学会第 25 回年次大会 発表論文集 , 2019.

[7]P. Zhou and N. El-Gohary, “Ontology-based automated information extraction from building energy conservation codes,” Automation in Construction, vol. 74, pp. 103-117 (2017)

（2020 年 11 月 10 日）

著 者 紹 介

著者：出町　和之
所属：東京大学大学院工学系研究科原子力専攻　准教授
専門分野:原子力保全工学、核セキュリティ工学、医用画像工学